ON THE TWO WEIGHT HILBERT TRANSFORM INEQUALITY

MICHAEL T. LACEY

ABSTRACT. Let σ and w be locally finite positive Borel measures on $\mathbb{R}$ which do not share a common point mass. Assume that the pair of weights satisfy a Poisson A_2 condition, and satisfy the testing conditions below, for the Hilbert transform H,

$$\int_I H(\sigma \mathbf{1}_I)^2 \, dw \lesssim \sigma(I)\,, \qquad \int_I H(w \mathbf{1}_I)^2 \, d\sigma \lesssim w(I)\,,$$

with constants independent of the choice of interval I. Then $H(\sigma \cdot)$ maps $L^2(\sigma)$ to $L^2(w)$, verifying a conjecture of Nazarov–Treil–Volberg. The proof uses basic tools of non-homogeneous analysis with two components particular to the Hilbert transform. The first is a global to local reduction, a consequence of prior work of Lacey-Sawyer-Shen-Uriate-Tuero. The second, an analysis of the local part, is the contribution of this paper.

CONTENTS

1. INTRODUCTION

Given weights (i.e. locally bounded positive Borel measures) σ and w on the real line $\mathbb{R}$, we consider the following *two weight norm inequality for the Hilbert transform,*

$$\int_{\mathbb{R}} |H_\epsilon(f\sigma)|^2 \, w(dx) \le \mathcal{N}^2 \int_{\mathbb{R}} |f|^2 \, \sigma(dx), \qquad f \in L^2(\sigma), \tag{1.1}$$

where $\mathcal{N}$ is the best constant in the inequality, uniform over all $0 < \epsilon < 1$, which define a standard truncation of the Hilbert transform applied to a signed locally finite measure ν

$$H_\epsilon \nu(x) := \int_{\epsilon < |x-y| < \epsilon^{-1}} \frac{\nu(dy)}{y - x}\,.$$

Research supported in part by grant NSF-DMS 0968499, and a grant from the Simons Foundation (#229596 to Michael Lacey). The author benefited from the research program Operator Related Function Theory and Time-Frequency Analysis at the Centre for Advanced Study at the Norwegian Academy of Science and Letters in Oslo during 2012—2013.

We insist upon this formulation as the principal value need not exist in the generality that we are interested in. Below, however, we systematically suppress the uniformity over ϵ above, writing just H for H_ϵ, with it understood that all estimates are independent of $0 < \epsilon < 1$.

A question of fundamental importance is establishing characterizations of the inequality above. In this paper we answer a conjecture of Nazarov–Treil–Volberg [18, 31], and sharpen a prior characterization of Lacey-Sawyer-Shen-Uriate-Tuero [11].

1.2. **Theorem.** *Let σ, w be two weights which do not share a common point mass. The inequality (1.1) holds if and only if the pair of weights σ, w satisfy these inequalities uniformly over all intervals I, and in their dual formulation. (The dual inequalities are obtained by interchanging the roles of w and σ.)*

$$\int_{\mathbb{R}} \frac{|I|}{(\operatorname{dist}(x, I) + |I|)^2} \sigma(dy) \cdot \frac{w(I)}{|I|} \le \mathcal{A}_2 \,, \tag{1.3}$$

$$\int_I H(\sigma \mathbf{1}_I)^2 \, w(dx) \le \mathcal{T}^2 \sigma(I) \,. \tag{1.4}$$

Taking $\mathcal{A}_2$ and $\mathcal{T}$ be the best constants of the inequalities above, there holds $\mathcal{N} \simeq \mathcal{A}_2^{1/2} + \mathcal{T}$.

This is an extension of the T1 Theorem of David-Journé [5], to a setting in which the transformation is fixed to be just a single operator, the Hilbert transform, but the weights are arbitrary. It is a further refinement of the real-variable characterization obtained by the author with Sawyer-Shen-Uriate-Tuero [11]. Indeed, the latter was the first such theorem for a continuous singular integral in the non-homogeneous setting pioneered in creative work of Nazarov–Treil–Volberg [15–18].

Note that the first condition is an extension of the typical A_2 condition to a 'half-Poisson' setting, which is known to be necessary. The second condition (1.4) is called an 'interval testing condition,' and is obviously necessary. Thus, the content of the Theorem is the sufficiency of the A_2 and testing conditions for the norm inequality.

If the pair of measures share a common point mass, $\sigma(\{x\}) \cdot w(\{x\}) > 0$ for some $x \in \mathbb{R}$, then the $\mathcal{A}_2$ condition is trivially false. In this case, one must change this condition, but then also check that the entire proof goes through, namely that there holds the energy inequality of [8], the functional energy inequality of [11], and that the argument of this paper goes through. This will be investigated elsewhere.

The study of two weight inequalities for individual operators was initiated by Eric Sawyer, with the maximal function [28], and later the fractional integrals [29]. Indeed, the latter characterization was precisely in the T1 language, and importantly for this paper, that paper also proved the two weight inequality for the Poisson integral, a fact used in [11, §7], delivering an inequality that is fundamental to the proof of the theorem above. The similarities between Sawyer's results and the T1 theorem were brought to the fore with the work on Nazarov–Treil–Volberg on non-homogeneous harmonic analysis already cited.

The question of the two weight inequality was raised by Muckenhoupt and Wheeden [13] in 1976. Sarason recognized the same problem in his deep work [27], in which he constructed examples individually unbounded Toeplitz operators whose composition was bounded. In [26],

Sarason raised the question: 'Characterize those pairs of outer functions g, h in H^2 of the unit disk such that the operator $T_g T_{\overline{h}}$ is bounded on H^2.' This is equivalent to the boundedness of the Riesz projection from $L^2(|g|^{-2})$ to $H^2(|f|^2)$, see [4, §5], and from here, the two weight question for the Hilbert transform gained wider attention. This same note of Sarason sketches a complex-variable argument of Treil proving the necessity of the A_2 condition.[1] Sarason wrote that it was 'tempting to conjecture' that the Poisson A_2 condition was sufficient for the norm inequality. That the simple A_2 condition is not sufficient is straight forward [13], but that the full Poisson A_2 condition is not sufficient is a deeper fact, proved by Nazarov [14]. This was an important step in formulating the conjecture solved in this paper.

The two weight inequality is also essentially equivalent to the question of characterizing those measures μ on the unit disk, for which a model space K_θ, θ inner, embeds into $L^2(\mathbb{D}, \mu)$. A question arose there, that could be understood as asking if the Poisson A_2 condition, and just one set of testing inequalities could be sufficient for the two weight inequality. This was disproved by Nazarov-Volberg [20]. The theory of Clark measures is highly relevant here, and in particular, there is no restriction on the class of measures that can arise as Clark measures, forcing one to consider arbitrary weights in this context. See [24] for more information. The Theorem of this paper also has applications to the theory of de Branges spaces [1], and to the spectral theory of (rank one perturbations of) normal operators [21].

Sufficient conditions for the boundedness of the composition of Toeplitz operators were given by Dechao Zheng, [32].[2] This particular direction leads to the so-called 'bump conditions' which remains an active research direction, with a somewhat different focus than ours.

In 2005, Nazarov–Treil–Volberg [18] created a method to prove two weight inequalities for Calderón–Zygmund operators for general measures, a component of their program of developing a non-homogeneous harmonic analysis. This innovative approach, incorporating the fundamental technique of random dyadic grids, and weight adapted martingale transform methods that have been integral to all subsequent approaches to this question, was strong enough to prove a certain variant of our main theorem for the triple of operators H_σ, M_σ and M_w, where M is the maximal function, see [18, Thm. 2.1]. This result was a notable success. Importantly, the argument proceeded by assuming that the pair of weights satisfied a supplemental pair of conditions, the so-called *pivotal conditions*, [18, §7.2]. This method of proof also had interesting implications [22] for the so-called A_2 theorem, solved by Hytönen [7].

The individual operator two weight problem is of course highly specific to the operator in question—in particular, there is no implication between the maximal function and the Hilbert transform in this context. Likewise, there are no weak-type estimates, Calderón–Zygmund decomposition, or interpolation available. While the two weight theory is largely complete for positive operators, the non-positive case is much harder. For certain kinds of dyadic operators, positive or non-positive, there is an elegant characterization in [19]. And, for well-behaved measures, one can

[1]Muckenhoupt-Wheeden used real-variable methods to show that the necessity of the half-Poisson A_2 condition. There is a complex-variable proof of the necessity of the full Poisson A_2 condition in [18, §3]. Also, the paper [13] includes results and conjectures in the L^p setting.

[2]In the language of this paper, the assumption is that w and σ have a density, and the Poisson A_2 condition is assumed, with a power bigger than one imposed on the densities. See [32, §6].

frequently reduce Calderón–Zygmund operators to dyadic ones. For the Hilbert transform, this can be done with the remarkable observation of Petermichl [23]; an extension of this to arbitrary Calderón–Zygmund operators is one of the important observations of Hytönen [7].

These dyadic methods obscure an essential aspect of the Hilbert transform, the last remnant of positivity: The kernel $\frac{1}{y-x}$ has a positive derivative in x. This is the main observation in the proof of necessity of the *energy inequality* (2.4), an essential strengthening of the Nazarov–Treil–Volberg pivotal conditions [18, §7.3]. The energy inequality is necessary from the A_2 and testing inequality. That is, it and it's further extensions are free to be used in the proof of the main theorem. Indeed, they must be used. The apparent difficulty in deriving the energy inequality from the dyadic models of the Hilbert transform are an important obstacle to dyadic approaches to our result.

The paper [8] by the author, Sawyer and Uriate-Tuero, proves the energy inequality. Using conditions which in a certain sense interpolate between the energy and pivotal conditions, sharper sufficient conditions were given for the the two weight inequality. These sharper conditions permit one to construct an example [8, §7] of a pair of weights which satisfy the two weight norm inequality, but fail the pivotal condition. In a certain sense, it is the best example known, in that simple modifications give alternate derivations of other counterexamples, including the examples of [14, 20], as well as the example in [25].[3]

Still this example gives only the barest hints of the inherent difficulties in the proof of our main theorem. The paper [10] introduced the natural Calderón–Zygmund stopping data into the subject, essential to the subsequent developments. There was an important breakthrough in a previous paper of the author, Sawyer-Shen-Uriate-Tuero [11] obtained an unconditional characterization of the two weight inequality. The characterization is in terms of the A_2 condition as in (1.3), but the testing conditions (1.4) were strengthened to $\int_I H(\sigma\mathbf{1}_E)^2 \; dw \lesssim \sigma(I)$ for all intervals and all Borel measurable subsets $E \subset I$, as well as the dual condition holding. There was no prior real-variable characterization known, whereas a complex-variable characterization, a variant of the Helson-Szegö theorem was established by Cotlar-Sadosky [3].

The proof of the main theorem, as mentioned uses the random grids and weight adapted martingale differences that are basic to the non-homogeneous theory. Then, aside from more routine considerations that are common to many proofs of T1 type theorems, the proof naturally splits into two parts. The first part is the reduction of the global L^2 inequality to one of a local nature. The essence of this part of the proof was found in a prior work of the author with Sawyer-Shen-Uriate-Tuero [11], which we recall in §3. This part depends upon a subtle multiscale extension of the energy inequality, one that itself is close to being stated in intrinsic form. It is proved in [11, §7] by appealing to Sawyer's two weight inequality for the Poisson integral [29].

After that, there is the control of the local part, which is largely contained in §4, a section devoted to the analysis of the so-called stopping form, with a highly non-intrinsic formulation. The stopping form is familiar to experts in the T1 theorem, but in all other settings, it is essentially an error term, expediently handled by some standard off-diagonal estimates. Any of these classical lines of reasoning will fail in the current setting. Instead, we construct a proof with a subtle

[3]There is a pair of weight σ, w with M_σ and M_w bounded, but H_σ not bounded.

recursion, one analogous to proofs of the Carleson theorem on the pointwise convergence of Fourier series [2, 6, 12]. It is the main novelty of this paper.

It is a pleasure to acknowledge the many conversations about this question that I have had with Ignacio Uriate-Tuero, Eric Sawyer, and Chun-Yun Shen.

2. Preliminaries

2.1. **Dyadic Grids.** A collection of intervals $\mathcal{G}$ is a *grid* if for all $G, G' \in \mathcal{G}$, we have $G \cap G' \in \{\emptyset, G, G'\}$. By a *dyadic grid* we mean a grid $\mathcal{D}$ of intervals of $\mathbb{R}$ such that for each interval $I \in \mathcal{D}$, the subcollection $\{I' \in \mathcal{D} \,:\, |I'| = |I|\}$ partitions $\mathbb{R}$, aside from endpoints of the intervals. In addition, the left and right halves of I, denoted by $I_\pm$, are also in $\mathcal{D}$.

For $I \in \mathcal{D}$, the left and right halves $I_\pm$ are referred to as the *children* of I. We denote by $\pi_{\mathcal{D}}(I)$ the unique interval in $\mathcal{D}$ having I as a child, and we refer to $\pi_{\mathcal{D}}(I)$ as the $\mathcal{D}$-parent of I.

We will work with subsets $\mathcal{F} \subset \mathcal{D}$. We say that I has $\mathcal{F}$ parent $\pi_{\mathcal{F}} I = F$ if $F \in \mathcal{F}$ is the minimal element of $\mathcal{F}$ that contains I. We also set $\pi^1_{\mathcal{F}} I := \pi_{\mathcal{F}} I$, and inductively set $\pi^{t+1}_{\mathcal{F}} I$ to be the minimal element of $\mathcal{F}$ that strictly contains $\pi^t_{\mathcal{F}} I$. The *$\mathcal{F}$-children* of $F \in \mathcal{F}$ are the maximal $F' \in \mathcal{F}$ which are strictly contained in F.

2.2. **Haar Functions.** Let σ be a weight on $\mathbb{R}$, one that does not assign positive mass to any endpoint of a dyadic grid $\mathcal{D}$. If $I \in \mathcal{D}$ is such that σ assigns non-zero weight to both children of I, the associated Haar function is chosen to have a non-negative inner product with the independent variable, $\langle x, h^\sigma_I(x) \rangle_\sigma \geq 0$, a convenient choice due to the central role of the energy inequality, (2.4).

$$h^\sigma_I := \sqrt{\frac{\sigma(I_-)\sigma(I_+)}{\sigma(I)}} \left(\frac{I_+}{\sigma(I_+)} - \frac{I_-}{\sigma(I_-)} \right). \tag{2.1}$$

In this definition, we are identifying an interval with its indicator function, and we will do so throughout the remainder of the paper. This is an $L^2(\sigma)$-normalized function, and has σ-integral zero. For any dyadic interval I_0, it holds that $\{\sigma(I_0)^{-1/2} I_0\} \cup \{h^\sigma_I : I \in \mathcal{D}, I \subset I_0\}$ is an orthonormal basis for $L^2(I_0, \sigma)$. We will use the notation $L^2_0(I_0, \sigma)$ for the subspace of $L^2(I_0, \sigma)$ of functions with mean zero. It has orthonormal basis $\{h^\sigma_I \,:\, I \in \mathcal{D}, I \subset I_0\}$.

We will use the notations $\hat{f}(I) = \langle f, h^\sigma_I \rangle_\sigma$, as well as

$$\Delta^\sigma_I f = \langle f, h^\sigma_I \rangle_\sigma h^\sigma_I = I_+ \mathbb{E}^\sigma_{I_+} f + I_- \mathbb{E}^\sigma_{I_-} f - I \mathbb{E}^\sigma_I f .$$

The second equality is the familiar martingale difference equality, and so we will refer to $\Delta^\sigma_I f$ as a martingale difference. It implies the familiar telescoping identity $\mathbb{E}^\sigma_J f = \sum_{I \,:\, I \supsetneq J} \mathbb{E}^\sigma_J \Delta^\sigma_I f$.

The *Haar support* of a function $f \in L^2(\sigma)$ is the collection $\{I \,:\, \hat{f}(I) \neq 0\}$.

2.3. **Good-Bad Decomposition.** Since the works of Nazarov–Treil–Volberg [15–17], the use of random dyadic grids is a foundational technique in the settings in which the measures are non-doubling. Our uses of them employs only standard and well-known facts.

With a choice of dyadic grid $\mathcal{D}$ understood, we say that $J \in \mathcal{D}$ is (ϵ, r)-*good* if and only if for all intervals $I \in \mathcal{D}$ with $|I| \geq 2^{r-1}|J|$, the distance from J to the boundary of *either child* of I is at least $|J|^{\epsilon}|I|^{1-\epsilon}$.

For $f \in L^2(\sigma)$ we set $P^{\sigma}_{\text{good}} f = \sum_{\substack{I \in \mathcal{D} \\ I \text{ is } (\epsilon, r)\text{-good}}} \Delta^{\sigma}_I f$. The projection $P^{w}_{\text{good}} g$ is defined similarly. With $0 < \epsilon < 1$ fixed, one will need to take $r > \epsilon^{-1}$, and any sufficiently large finite value suffices. The property of intervals being (ϵ, r)-good is essential, and highlighted when the property is being used.

2.4. **Energy, Monotonicity, and Poisson.** We collect results specific to the Hilbert transform; see [8, §2] and [11, §5] for more details. Throughout the paper, we use this definition of the Poisson integral of weight σ over interval I.

$$P(\sigma, I) := \int_{\mathbb{R}} \frac{|I|}{(\operatorname{dist}(x, I) + |I|)^2} \sigma(dy) \,.$$

Frequently, σ has a further restriction on its support, clearly indicated in the notation.

2.2. **Lemma** (Monotonicity Property)**.** *Suppose that ν is a signed measure, and μ is a positive measure with $\mu \geq |\nu|$, both supported outside an interval $I \in \mathcal{D}$. Then, for good $J \subset I$ and $2^{r+1}|J| \leq |I|$, and function $g \in L^2_0(J, w)$, it holds that*

$$|\langle H\nu, g\rangle_w| \leq \langle H\mu, \overline{g}\rangle_w \approx P(\mu, J)\Big\langle \frac{x}{|J|}, \overline{g}\Big\rangle_w. \tag{2.3}$$

Here, $\overline{g} = \sum_{J'} |\widehat{g}(J')| h^w_{J'}$, is a Haar multiplier applied to g.

The concept of *energy* is fundamental to the subject. For interval I, define

$$E(w, I)^2 := \frac{1}{|I|} \sum_{J \,:\, J \subset I} \langle x, h^w_J\rangle^2_w$$

Now, consider the *energy constant*, the smallest constant $\mathcal{E}$ such that this condition holds, as presented or in its dual formulation. For all intervals I_0, all partitions $\mathcal{P}$ of I_0, it holds that

$$\sum_{I \in \mathcal{P}} P(\sigma I_0, I)^2 E(w, I)^2 w(I) \leq \mathcal{E}^2 \sigma(I_0) \,. \tag{2.4}$$

2.5. **Lemma.** *[Energy Inequality] There holds $\mathcal{E} \lesssim \mathcal{A}_2^{1/2} + \mathcal{T} = \mathcal{H}$.*

For a proof, see [8, Proposition 2.11].

2.6. *Remark.* One should note that for interval $J \subset I_0$, and interval $J' \subset J$, there holds

$$\langle H_{\sigma}(I_0 - J), h^w_{J'}\rangle_w \gtrsim P(\sigma(I_0 - J), J')\Big\langle \frac{x}{|J'|}, h^w_{J'}\Big\rangle_w \,.$$

And, while the inequality is strict in general, we can *reverse it* if J' is good and $J' \Subset J$. This distinction is basic to the subject, and drives some of the case analysis in the proof of Lemma 4.6.

2.7. *Remark.* The influential pivotal condition of Nazarov–Treil–Volberg [18] is obtained from the energy condition by setting $E(w, I) \equiv 1$. Namely, it is the assumption on the pair of weights that there is a finite positive constant $\mathcal{P}$ such that for all intervals I_0, all partitions $\mathcal{P}$ of I_0, it holds that

$$\sum_{I \in \mathcal{P}} P(\sigma I_0, I)^2 w(I) \le \mathcal{P}^2 \sigma(I_0) \,.$$

And, the dual inequalities also hold. Note that $P(\sigma I_0, I) \lesssim \inf_{x \in I} M(\sigma I_0)(x)$, so that boundedness of the maximal functions $M_\sigma : L^2(\sigma) \to L^2(w)$ and $M_w : L^2(w) \to L^2(\sigma)$ imply the pivotal condition. Though they wrote that the pivotal conditions 'might turn out to be necessary', this was disproved in [8].

3. The Global to Local Reduction

The first half of the proof follows from the techniques of [11], though that paper does not prove a result in the form that we need it. The goal is the reduction to the local estimate, (3.14), at the end of this section.

Our aim is to prove

$$|\langle H_\sigma f, g \rangle_w| \lesssim \mathcal{H} \|f\|_\sigma \|g\|_w \,, \tag{3.1}$$

where here and throughout $\mathcal{H} := \mathcal{A}_2^{1/2} + \mathcal{T}$. And, as methods are of necessity focused on L^2, we systematically abbreviate $\|f\|_{L^2(\sigma)}$ to $\|f\|_\sigma$.

The functions $f \in L^2(\sigma)$, and $g \in L^2(w)$ are expanded with respect to the Haar basis with respect to a fixed dyadic grid $\mathcal{D}$, and adapted to the weight in question.

A reduction, using randomized dyadic grids, allows one the extraordinarily useful reduction in the next Lemma. This is a well-known reduction, due to Nazarov–Treil–Volberg, explained in full detail in the current setting, in [18, §4].

3.2. **Lemma.** *For all sufficiently small ϵ, and sufficiently large r, this holds. Suppose that for any dyadic grid $\mathcal{D}$, such that no endpoint of an interval $I \in \mathcal{D}$ is a point mass for σ or w,*[4] *there holds*

$$|\langle H_\sigma P^\sigma_{\text{good}} f, P^w_{\text{good}} g \rangle_w| \lesssim \mathcal{H} \|f\|_\sigma \|g\|_w \,.$$

Then, the same inequality holds without the projections P^σ_{good}, and P^w_{good}, namely (3.1) holds.

That is, the bilinear form only needs to be controlled for (ϵ, r)-*good functions* f *and* g*, goodness being defined with respect to a fixed dyadic grid.* Suppressing the notation, we write 'good' for '(ϵ, r)-good,' and it is always assumed that the dyadic grid $\mathcal{D}$ is fixed, and only good intervals are in the Haar support of f and g, though is also suppressed in the notation. We clearly remark on goodness when the property is used.

It is sufficient to assume that f and g are supported on an interval I_0; by trivial use of the interval testing condition, we can further assume that f and g are of integral zero in their respective

[4] This set of dyadic grids that fail this condition have probability zero in standard constructions of the random dyadic grids.

spaces. Thus, f is in the linear span of (good) Haar functions h^{σ}_I for $I \subset I_0$, and similarly for g, and

$$\langle H_\sigma f, g\rangle_w = \sum_{I,J\,:\,I,J\subset I_0} \langle H_\sigma \Delta^{\sigma}_I f, \Delta^{w}_J g\rangle_w \,.$$

The double sum is broken into different summands. Many of the resulting cases are elementary, and we summarize these estimates as follows. Define the bilinear form

$$B^{\textup{above}}(f,g) := \sum_{I\,:\,I\subset I_0} \sum_{J\,:\,J\Subset I} \mathbb{E}^{\sigma}_J \Delta^{\sigma}_I f \cdot \langle H_\sigma I_J, \Delta^{w}_J g\rangle_w$$

where here and throughout, $J \Subset I$ means $J \subset I$ and $2^{r+1}|J| \le |I|$. In addition, the argument of the Hilbert transform, I_J, is the child of I that contains J, so that $\Delta^{\sigma}_I f$ is constant on I_J, and $\mathbb{E}^{\sigma}_J \Delta^{\sigma}_I f = \mathbb{E}^{\sigma}_{I_J} \Delta^{\sigma}_I f$. Define $B^{\textup{below}}(f,g)$ in the dual fashion.

3.3. **Lemma.** *There holds*

$$\Bigl\lvert \langle H_\sigma f, g\rangle_w - B^{\textup{above}}(f,g) - B^{\textup{below}}(f,g)\Bigr\rvert \lesssim \mathcal{H} \lVert f\rVert_\sigma \lVert g\rVert_w \,.$$

This is a common reduction in a proof of a T1 theorem, and in the current context, it only requires goodness of intervals and the A_2 condition. For a proof, one can consult [18, 31]. The Lemma is specifically phrased and proved in this way in [10, §8].

Thus, the main technical result is as below; it immediately supplies our main theorem.

3.4. **Theorem.** *There holds*

$$|B^{\textup{above}}(f,g)| \lesssim \mathcal{H} \lVert f\rVert_\sigma \lVert g\rVert_w \,.$$

The same inequality holds for the dual form $B^{\textup{below}}(f,g)$.

3.5. *Remark.* We emphasize that partial information about $B^{\textup{above}}(f,g)$ may not yield any substantive information about the bilinear form $\langle H_\sigma f, g\rangle_w$.[5] Nevertheless, [11] proved a characterization of the two weight inequality–using the a different tool, the so-called *parallel corona.*

In the remainder of this section, we recall techniques from [11] that permit reduction of the global Theorem 3.4 to a localized setting in which the function f is more structured in that it has bounded averages on a fixed interval, and the pair of function f, g are more structured in that their Haar supports avoid intervals that strongly violate the energy inequality.

3.6. **Definition.** Given any interval I_0, define $\mathcal{F}_{\textup{energy}}(I_0)$ to be the maximal subintervals $I \subsetneq I_0$ such that

$$P(\sigma I_0, J)^2 E(w,J)^2 w(J) > 10\mathcal{E}^2 \sigma(I) \,.$$

Here, $\mathcal{E}$ is the constant in (2.4), and it holds that $\mathcal{E} \lesssim \mathcal{H}$. There holds $\sigma(\cup\{F : F \in \mathcal{F}(I_0)\}) \le \frac{1}{10}\sigma(I_0)$, by the energy inequality.

[5] To illustrate, using the techniques of [11], one can show that the norm of the bilinear form $B^{\textup{above}}(f,g)$ is dominated by $\mathcal{H}+\mathcal{B}_\infty$, where the latter constant is the best constant in $|B^{\textup{above}}(f,g)| \le \mathcal{B}_\infty \sigma(I_0)^{1/2} \lVert g\rVert_w$, where $|f| \le I_0$. This highly non-trivial fact has no implication for the full bilinear form $\langle H_\sigma f, g\rangle_w$.

We make the following construction for an $f \in L^2_0(I_0, \sigma)$, the subspace of $L^2(I_0, \sigma)$ of functions of mean zero. Add I_0 to $\mathcal{F}$, and set $\alpha_f(I_0) := \mathbb{E}^\sigma_{I_0}|f|$. In the inductive stage, if $F \in \mathcal{F}$ is minimal, add to $\mathcal{F}$ those maximal descendants F' of F such that $F' \in \mathcal{F}_{\text{energy}}(F)$ or $\mathbb{E}^\sigma_{F'}|f| \geq 10\alpha_f(F)$. Then define

$$\alpha_f(F') := \begin{cases} \alpha_f(F) & \mathbb{E}^\sigma_{F'}|f| < 10\alpha_f(F) \\ \mathbb{E}^\sigma_{F'}|f| & \text{otherwise} \end{cases}$$

If there are no such intervals F', the construction stops. We refer to $\mathcal{F}$ and $\alpha_f(\cdot)$ as *Calderón–Zygmund stopping data for* f, following the terminology of [10, Def 3.5], [11, Def 3.4]. Their key properties are collected here.

3.7. **Lemma.** *For $\mathcal{F}$ and $\alpha_f(\cdot)$ as defined above, there holds*

(1) I_0 *is the maximal element of $\mathcal{F}$.*
(2) *For all $I \in \mathcal{D}$, $I \subset I_0$, we have $\mathbb{E}^\sigma_I|f| \leq 10\alpha_f(\pi_{\mathcal{F}} I)$.*
(3) *α_f is monotonic: If $F, F' \in \mathcal{F}$ and $F \subset F'$ then $\alpha_f(F) \geq \alpha_f(F')$.*
(4) *The collection $\mathcal{F}$ is σ-Carleson in that*

$$\sum_{F \in \mathcal{F}\colon\, F \subset S} \sigma(F) \leq 2\sigma(S), \qquad S \in \mathcal{D}. \tag{3.8}$$

(5) *We have the inequality*

$$\Big\| \sum_{F \in \mathcal{F}} \alpha_f(F) \cdot F \Big\|_\sigma \lesssim \|f\|_\sigma \,. \tag{3.9}$$

Proof. The first three properties are immediate from the construction. The fourth, the σ-Carleson property is seen this way. It suffices to check the property for $S \in \mathcal{F}$. Now, the $\mathcal{F}$-children can be in $\mathcal{F}_{\text{energy}}(S)$, which satisfy

$$\sum_{F' \in \mathcal{F}_{\text{energy}}(S)} \sigma(F') \leq \tfrac{1}{10}\sigma(S)\,.$$

Or, they satisfy $\mathbb{E}^\sigma_{F'}|f| \geq 10\mathbb{E}^\sigma_S|f|$, but these intervals satisfy the same estimate. Hence, (3.8) holds.

For the final property, let $\mathcal{G} \subset \mathcal{F}$ be the subset at which the stopping values change: If $F \in \mathcal{F} - \mathcal{G}$, and G is the $\mathcal{G}$-parent of F, then $\alpha_f(F) = \alpha_f(G)$. Set

$$\Phi_G := \sum_{F \in \mathcal{F}\,:\, \pi_{\mathcal{G}} F = G} F\,.$$

Define $G_k := \{\Phi_G \geq 2^k\}$, for $k = 0, 1, \ldots$. The σ-Carleson property implies integrability of all orders in σ-measure of Φ_G. Using the third moment, we have $\sigma(G_k) \lesssim 2^{-3k}\sigma(G)$. Then, estimate

$$\Big\| \sum_{F \in \mathcal{F}} \alpha_f(F) \cdot F \Big\|_\sigma^2 = \Big\| \sum_{G \in \mathcal{G}} \alpha_f(G) \Phi_G \Big\|_\sigma^2$$

$$\begin{aligned}
&\le \Big\| \sum_{k=0}^{\infty} (k+1)^{+1-1} \sum_{G\in\mathcal{G}} \alpha_f(G) 2^k \mathbf{1}_{G_k} \Big\|_\sigma^2 \\
&\overset{*}{\lesssim} \sum_{k=0}^{\infty} (k+1)^2 \Big\| \sum_{G\in\mathcal{G}} \alpha_f(G) 2^k \mathbf{1}_{G_k}(x) \Big\|_\sigma^2 \\
&\overset{**}{\lesssim} \sum_{k=0}^{\infty} (k+1)^2 \sum_{G\in\mathcal{G}} \alpha_f(G)^2 2^{2k} \sigma(G_k) \\
&\lesssim \sum_{G\in\mathcal{G}} \alpha_f(G)^2 \sigma(G) \lesssim \|Mf\|_\sigma^2 \lesssim \|f\|_\sigma^2 .
\end{aligned}$$

Note that we have used Cauchy–Schwarz in k at the step marked by an $*$. In the step marked with $**$, for each point x, the non-zero summands are a (super)-geometric sequence of scalars, so the square can be moved inside the sum. Finally, we use the estimate on the σ-measure of G_k, and compare to the maximal function Mf to complete the estimate. □

We will use the notation

$$P_F^\sigma f := \sum_{I\in\mathcal{D}\,:\,\pi_{\mathcal{F}} I = F} \Delta_I^\sigma f , \qquad F \in \mathcal{F} .$$

and similarly for Q_F^w. The inequality (3.9) allows us to estimate

$$\begin{aligned}
&\sum_{F\in\mathcal{F}} \{\alpha_f(F)\sigma(F)^{1/2} + \|P_F^\sigma f\|_\sigma\} \|Q_F^w g\|_w \\
&\qquad \le \Big[\sum_{F\in\mathcal{F}} \{\alpha_f(F)^2\sigma(F) + \|P_F^\sigma f\|_\sigma^2\} \times \sum_{F\in\mathcal{F}} \|Q_F^w g\|_w^2 \Big]^{1/2} \lesssim \|f\|_\sigma \|g\|_w .
\end{aligned} \tag{3.10}$$

We will refer to as the *quasi-orthogonality* argument. It is very useful.

The Theorem below is the essence of the reduction from a global to local estimate in our proof.

3.11. **Theorem.** *[Global to Local Reduction] There holds*

$$\begin{aligned}
&\Big| B^{\mathrm{above}}(f,g) - B^{\mathrm{above}}_{\mathcal{F}}(f,g) \Big| \lesssim \mathcal{H} \|f\|_\sigma \|g\|_w , \\
&\text{where} \quad B^{\mathrm{above}}_{\mathcal{F}}(f,g) := \sum_{F\in\mathcal{F}} B^{\mathrm{above}}(P_F^\sigma f, Q_F^w g) .
\end{aligned}$$

A reduction of this type is a familiar aspect of many proofs of a T1 theorem, proved by exploiting standard off-diagonal estimates for Calderón–Zygmund kernels. It is one of the contributions of [18] to point out that such arguments are far more sophisticated in the two weight setting, and [11] showed that, with Calderón–Zygmund stopping data, the reduction can be made assuming the A_2 and testing hypotheses.

Proof. This is an immediate corollary to [11, Theorem 6.6], but we include enough detail here so that the reader need only examine the proof of the functional energy inequality in [11, §7]. The

latter inequality is a sophisticated extension of the energy inequality, one of the key innovations of that paper. It's proof is self-contained, except for the two weight inequality for the Poisson integral, proved by Sawyer [29, Thm 2].[6] We have

$$\mathrm{B}^{\mathrm{above}}(f,g)=\sum_{F'\in\mathcal{F}}\sum_{F\in\mathcal{F}\,:\,F\subset F'}\mathrm{B}^{\mathrm{above}}(\mathrm{P}^{\sigma}_{F'}f,\mathrm{Q}^{w}_{F}g)\,.$$

The form $\mathrm{B}^{\mathrm{above}}_{\mathcal{F}}(f,g)$ is the case of $F=F'$ in the double sum above, hence we should bound

$$\sum_{F\in\mathcal{F}}\mathrm{B}^{\mathrm{above}}(\Phi_F f,\mathrm{Q}^{w}_{F}g)=\sum_{F\in\mathcal{F}}\langle \mathrm{H}_{\sigma}\Phi_F,\mathrm{Q}^{w}_{F}g\rangle_{w}\,, \tag{3.12}$$
$$\text{where}\quad \Phi_F:=\sum_{F'\in\mathcal{F}\,:\,F'\supsetneq F}\mathrm{P}^{\sigma}_{F'}f\,.$$

In the language of [11, Def 7.1], the sequence of functions $\{\mathrm{Q}^{w}_{F}g\ :\ F\in\mathcal{F}\}$ is $\mathcal{F}_{\subset}$-adapted, a key component of the functional energy inequality. In particular, from [11, Cor 7.5], there follows

$$\Bigl|\sum_{F\in\mathcal{F}}\sum_{I\,:\,I\supsetneq F}\mathbb{E}^{\sigma}_{I_F}\Delta^{\sigma}_{I}f\cdot\langle \mathrm{H}_{\sigma}I_F,\mathrm{Q}^{w}_{F}g\rangle_{w}\Bigr|\lesssim\mathcal{H}\|f\|_{\sigma}\Bigl[\sum_{F\in\mathcal{F}}\|\mathrm{Q}^{w}_{F}g\|^2_{w}\Bigr]^{1/2}=\mathcal{H}\|f\|_{\sigma}\|g\|_{w}\,.$$

By inspection, the sum on the left equals the expression in (3.12), so the proof is complete, aside from the functional energy inequality. □

It remains to control $\mathrm{B}^{\mathrm{above}}_{\mathcal{F}}(f,g)$. Keeping the quasi-orthogonality argument in mind, we see that appropriate control on the individual summands is enough to control it. To describe what has been done, one must note that the functions $\mathrm{P}^{\sigma}_{F}f$ *need not be bounded*. But, they have bounded averages, and both functions $\mathrm{P}^{\sigma}_{F}f$ and $\mathrm{Q}^{w}_{F}g$ are well-adapted to the pair of weights w,σ. This is formalized in the next definition.

3.13. **Definition.** Let I_0 be an interval, and let $\mathcal{S}$ be a collection of disjoint intervals contained in S. A function $f\in L^2_0(I_0,\sigma)$ is said to be *uniform (w.r.t. $\mathcal{S}$)* if these conditions are met:

1. Each energy stopping interval $F\in\mathcal{F}_{\mathrm{energy}}(I_0)$ is contained in some $S\in\mathcal{S}$.
2. The function f is constant on each interval $S\in\mathcal{S}$.
3. For any interval I which is not contained in any $S\in\mathcal{S}$, $\mathbb{E}^{\sigma}_{I}|f|\le 1$.

We will say that g is *adapted* to a function f uniform w.r.t. $\mathcal{S}$, if g is constant on each interval $S\in\mathcal{S}$. We will also say that g is *adapted to $\mathcal{S}$.*

Let us define what we mean by the *local estimate.* The constant $\mathcal{B}_{\mathrm{local}}$ is defined as the best constant in

$$|\mathrm{B}^{\mathrm{above}}(f,g)|\le\mathcal{B}_{\mathrm{local}}\{\sigma(I_0)^{1/2}+\|f\|_{\sigma}\}\|g\|_{w}\,, \tag{3.14}$$

where f,g are of mean zero on their respective spaces, supported on an interval I_0. Moreover, f is uniform, and g is adapted to f. The inequality above is homogeneous in g, but not f, since the term $\sigma(I_0)^{1/2}$ is motivated by the bounded averages property of f.

[6] A gap in the proof of the Poisson inequality at [29, Page 542] can be fixed as in [30] or [9, Lemma 4.10].

The reduction from global to local estimate is Theorem 3.11. The Lemma below, shows that it suffices to bound the local estimate.

3.15. **Lemma.** *There holds*

$$|\mathrm{B}^{\text{above}}(f, g)| \lesssim \{\mathcal{B}_{\text{local}} + \mathcal{H}\}\|f\|_\sigma \|g\|_w .$$

Proof. Let $\mathcal{F}$ and $\alpha_f(\cdot)$ be standard Calderón–Zygmund stopping data for f. By Theorem 3.11, it suffices to bound

$$\mathrm{B}^{\text{above}}_{\mathcal{F}}(f, g) = \sum_{F \in \mathcal{F}} \mathrm{B}^{\text{above}}(\mathrm{P}^\sigma_F f, \mathrm{Q}^w_F g)$$

For each $F \in \mathcal{F}$, let $\mathcal{S}_F$ be the $\mathcal{F}$-children of F. Observe that the function

$$(C\alpha_f(F))^{-1}\mathrm{P}^\sigma_F f \tag{3.16}$$

is uniform on F w.r.t. $\mathcal{S}_F$, for appropriate absolute constant C. Moreover, the function $\mathrm{Q}^w_F g$ does not have any interval J in its Haar support contained in an interval $S \in \mathcal{S}_F$. That is, it is adapted to the function in (3.16). Therefore, by assumption,

$$|\mathrm{B}^{\text{above}}(\mathrm{P}^\sigma_F f, \mathrm{Q}^w_F g)| \le \mathcal{B}_{\text{local}}\{\alpha_F(F)\sigma(F)^{1/2} + \|\mathrm{P}^\sigma_F f\|_\sigma\}\|\mathrm{Q}^w_F g\|_w .$$

The sum over $F \in \mathcal{F}$ of the right hand side is bounded by the quasi-orthogonality argument of (3.10). □

Thus, it remains to show that $\mathcal{B}_{\text{local}} \lesssim \mathcal{H}$. The following reduction in the local estimate is a routine appeal to the testing condition. Focusing on the argument of the Hilbert transform in (3.14), we write $I_J = I_0 - (I_0 - I_J)$. When the interval is I_0, and J is in the Haar support of g, notice that the scalar

$$\varepsilon_J := \sum_{I \,:\, J \Subset I \subset I_0} \mathbb{E}^\sigma_J \Delta^\sigma_I f$$

is bounded by one. Say that f is uniform w.r.t. $\mathcal{S}$, and let I^- be the minimal interval in the Haar support of f with $J \Subset I$. Since g is adapted to f, we cannot have I^-_J contained in an interval of $\mathcal{S}$, and so $|\mathbb{E}^\sigma_{I^-_J} f| \le 1$. By the telescoping identity for martingale differences,

$$\varepsilon_J = \sum_{I \,:\, I^- \subset I \subset I_0} \mathbb{E}^\sigma_{I_J} \Delta^\sigma_I f = \mathbb{E}^\sigma_{I^-_J} f ,$$

which is at most one in absolute value.

Therefore, we can write

$$\begin{aligned}
\Big| \sum_{I \,:\, I \subset I_0} \sum_{J \,:\, J \Subset I} \mathbb{E}^\sigma_J \Delta^\sigma_I f \cdot \langle \mathrm{H}_\sigma I_0, \Delta^w_J g \rangle \Big| &= \Big| \Big\langle \mathrm{H}_\sigma I_0, \sum_{J \,:\, J \Subset I_0} \varepsilon_J \Delta^w_J g \Big\rangle_w \Big| \\
&\le \mathcal{T}\sigma(I_0)^{1/2} \Big\| \sum_{J \,:\, J \Subset I_0} \varepsilon_J \Delta^w_J g \Big\|_w \\
&\le \mathcal{T}\sigma(I_0)^{1/2} \|g\|_w .
\end{aligned}$$

This uses only interval testing and orthogonality of the martingale differences, and it matches the first half of the right hand side of (3.14).

When the argument of the Hilbert transform is $I_0 - I_J$, this is the *stopping form*, the last component of the local part of the problem. The treatment of it, in the next section, is the main novelty of this paper.

4. The Stopping Form

Given an interval I_0, the stopping form is

$$(4.1)\qquad B^{\mathsf{stop}}_{I_0}(f,g) := \sum_{I\,:\,I\subset I_0}\ \sum_{J\,:\,J\Subset I_J} \mathbb{E}^{\sigma}_{I_J}\Delta^{\sigma}_{I} f\cdot \langle H_\sigma(I_0 - I_J), \Delta^{w}_{J} g\rangle_w\,.$$

We prove this for the stopping form, which completes the proof of the inequality $\mathcal{B}_{\mathsf{local}} \lesssim \mathcal{H}$, and so in view of Lemma 3.15, completes the proof of the main theorem of this paper. Note that the hypotheses on f and g are that they are adapted to energy stopping intervals. (Bounded averages on f are no longer required.)

4.2. **Lemma.** *Fix an interval* I_0*, and let* f *and* g *be be adapted to* $\mathcal{F}_{\mathsf{energy}}(I_0)$*. Then,*

$$|B^{\mathsf{stop}}_{I_0}(f,g)| \lesssim \mathcal{H}\|f\|_\sigma \|g\|_w\,.$$

The stopping form arises naturally in any proof of a T1 theorem using Haar or other bases. In the non-homogeneous case, or in the Tb setting, where (adapted) Haar functions are important tools, it frequently appears in more or less this form. Regardless of how it arises, the stopping form is treated as a error, in that it is bounded by some simple geometric series, obtaining decay as e. g. the ratio $|J|/|I|$ is held fixed. (See for instance [18, (7.16)].)

These sorts of arguments, however, implicitly require some additional hypotheses, such as the weights being mutually A_∞. Of course, the two weights above can be mutually singular. There is no *a priori* control of the stopping form in terms of simple parameters like $|J|/|I|$, even supplemented by additional pigeonholing of various parameters.

Our method is inspired by proofs of Carleson's Theorem on Fourier series [2, 6, 12], and has one particular precedent in the current setting, a much simpler bound for the stopping form in [11, §6.1].

4.1. **Admissible Pairs.**

A range of decompositions of the stopping form necessitate a somewhat heavy notation that we introduce here. The individual summands in the stopping form involve four distinct intervals, namely I_0, I, I_J, and J. The interval I_0 will not change in this argument, and the pair (I, J) determine I_J. Subsequent decompositions are easiest to phrase as actions on collections $\mathcal{Q}$ of pairs of intervals $Q = (Q_1, Q_2)$ with $Q_1 \Supset Q_2$. (The letter P is already taken for the Poisson integral.) And we consider the bilinear forms

$$B_{\mathcal{Q}}(f,g) := \sum_{Q\in\mathcal{Q}} \mathbb{E}^{\sigma}_{Q_2}\Delta^{\sigma}_{Q_1} f\cdot \langle H_\sigma(I_0 - (Q_1)_{Q_2}), \Delta^{w}_{Q_2} g\rangle_w\,.$$

We will have the standing assumption that for all collections $\mathcal{Q}$ that we consider are *admissible*.

4.3. **Definition.** A collection of pairs $\mathcal{Q}$ is *admissible* if it meets these criteria. For any $Q = (Q_1, Q_2) \in \mathcal{Q}$,

1. $Q_2 \Subset Q_1 \subset I_0$.
2. (convexity in Q_1) If $Q'' \in \mathcal{Q}$ with $Q''_2 = Q_2$ and $Q''_1 \subset I \subset Q_1$, then there is a $Q' \in \mathcal{Q}$ with $Q'_1 = I$ and $Q'_2 = Q_2$.

The first property is self-explanatory. The second property is convexity in Q_1, holding Q_2 fixed, which is used in the estimates on the stopping form which conclude the argument. A third property is described below.

We exclusively use the notation $\mathcal{Q}_k$, $k = 1, 2$ for the collection of intervals $\bigcup\{Q_k : Q \in \mathcal{Q}\}$, not counting multiplicity. Similarly, set $\tilde{\mathcal{Q}}_1 := \{(Q_1)_{Q_2} : Q \in \mathcal{Q}\}$, and $\tilde{Q}_1 := (Q_1)_{Q_2}$.

3. No interval $K \in \tilde{\mathcal{Q}}_1 \cup \mathcal{Q}_2$ is contained in an interval $S \in \mathcal{F}_{\textsf{energy}}(I_0)$.

The last requirement comes from the assumption that the functions f and g be adapted to $\mathcal{F}_{\textsf{energy}}(I_0)$. We will be appealing to different Hilbertian arguments below, so we prefer to make this an assumption about the pairs than the functions f, g.

The stopping form is obtained with the admissible collection of pairs given by

$$\mathcal{Q}_0 = \{(I, J) : J \Subset I, J \not\subset \cup\{S : \mathcal{S}\}\}. \tag{4.4}$$

In this definition $\mathcal{S}$ is the collection of subintervals of I_0 which f is uniform with respect to. There holds $B^{\textsf{stop}}_{I_0}(f, g) = B_{\mathcal{Q}_0}(f, g)$ for f, g adapted to $\mathcal{F}_{\textsf{energy}}(I_0)$.

There is a very important notion of the size of $\mathcal{Q}$.

$$\textsf{size}(\mathcal{Q})^2 := \sup_{K \in \tilde{\mathcal{Q}}_1 \cup \mathcal{Q}_2} \frac{\mathsf{P}(\sigma(I_0 - K), K)^2}{\sigma(K)|K|^2} \sum_{J \in \mathcal{Q}_2 : J \subset K} \langle x, h^w_J \rangle^2_w .$$

For admissible $\mathcal{Q}$, there holds $\textsf{size}(\mathcal{Q}) \lesssim \mathcal{H}$, as follows the property (3) in Definition 4.3, and Definition 3.6.

More definitions follow. Set the norm of the bilinear form $\mathcal{Q}$ to be the best constant in the inequality

$$|B_{\mathcal{Q}}(f, g)| \le \mathbf{B}_{\mathcal{Q}} \|f\|_\sigma \|g\|_w .$$

Thus, our goal is show that $\mathbf{B}_{\mathcal{Q}} \lesssim \textsf{size}(\mathcal{Q})$ for admissible $\mathcal{Q}$, but we will only be able to do this directly in the case that the pairs (Q_1, Q_2) are weakly decoupled.

Say that collections of pairs $\mathcal{Q}^j$, for $j \in \mathbb{N}$, are *mutually orthogonal* if on the one hand, the collections $(\mathcal{Q}^j)_2$ are pairwise disjoint, and on the other, that the collection $\widetilde{(\mathcal{Q}^j)}_1$ are pairwise disjoint. (The concept has to be different in the first and second coordinates of the pairs, due to the different role of the intervals Q_1 and Q_2.)

The meaning of mutual orthogonality is best expressed through the norm of the associated bilinear forms. Under the assumption that $B_{\mathcal{Q}} = \sum_{j \in \mathbb{N}} B_{\mathcal{Q}^j}$, and that the $\{\mathcal{Q}^j : j \in \mathbb{N}\}$ are mutually orthogonal, the following essential inequality holds.

$$\mathbf{B}_{\mathcal{Q}} \le \sqrt{2} \sup_{j \in \mathbb{N}} \mathbf{B}_{\mathcal{Q}^j} . \tag{4.5}$$

Indeed, for $j \in \mathbb{N}$, let Π_j^w be the projection onto the linear span of the Haar functions $\{h_J^w : J \in \mathcal{Q}_2^j\}$, and use a similar notation for Π_j^σ. We then have the two inequalities

$$\sum_{j \in \mathbb{N}} \lVert \Pi_j^w g \rVert_w^2 \le \lVert g \rVert_w^2 , \qquad \sum_{j \in \mathbb{N}} \lVert \Pi_j^\sigma f \rVert_\sigma^2 \le 2 \lVert f \rVert_\sigma^2 .$$

Note the factor of two on the second inequality. Therefore, we have

$$\begin{aligned} \lvert B_{\mathcal{Q}}(f,g) \rvert &\le \sum_{j \in \mathbb{N}} \lvert B_{\mathcal{Q}^j}(f,g) \rvert \\ &= \sum_{j \in \mathbb{N}} \lvert B_{\mathcal{Q}^j}(\Pi_j^\sigma f, \Pi_j^w g) \rvert \\ &\le \sum_{j \in \mathbb{N}} \mathbf{B}_{\mathcal{Q}^j} \lVert \Pi_j^\sigma f \rVert_\sigma \lVert \Pi_j^w g \rVert_w \le \sqrt{2} \sup_{j \in \mathbb{N}} \mathbf{B}_{\mathcal{Q}^j} \cdot \lVert f \rVert_\sigma \lVert g \rVert_w . \end{aligned}$$

This proves (4.5).

4.2. **The Recursive Argument.** This is the essence of the matter.

4.6. **Lemma.** *[Size Lemma] An admissible collection of pairs $\mathcal{Q}$ can be partitioned into collections $\mathcal{Q}^{\text{large}}$ and admissible $\mathcal{Q}_t^{\text{small}}$, for $t \in \mathbb{N}$ such that*

$$\begin{aligned} &\mathbf{B}_{\mathcal{Q}} \le C\,\mathsf{size}(\mathcal{Q}) + (1+\sqrt{2}) \sup_t \mathbf{B}_{\mathcal{Q}_t^{\text{small}}} , \\ &\textit{and} \quad \sup_{t \in \mathbb{N}} \mathsf{size}(\mathcal{Q}_t^{\text{small}}) \le \tfrac{1}{4} \mathsf{size}(\mathcal{Q}) . \end{aligned} \tag{4.7}$$

Here, $C > 0$ is an absolute constant.

The point of the lemma is that all of the constituent parts are better in some way, and that the right hand side of (4.7) involves a favorable supremum. We can quickly prove the main result of this section.

Proof of Lemma 4.2. The stopping form of this Lemma is of the form $B_{\mathcal{Q}}(f,g)$ for admissible choice of $\mathcal{Q}$, with $\mathsf{size}(\mathcal{Q}) \le C\mathcal{H}$, as we have noted in (4.4). Define

$$\zeta(\lambda) := \sup\{\mathbf{B}_{\mathcal{Q}} \,:\, \mathsf{size}(\mathcal{Q}) \le C\lambda\mathcal{H}\} , \qquad 0 < \lambda \le 1 ,$$

where $C > 0$ is a sufficiently large, but absolute constant, and the supremum is over admissible choices of $\mathcal{Q}$. We are free to assume that $\mathcal{Q}_1$ and $\mathcal{Q}_2$ are further constrained to be in some fixed, but large, collection of intervals $\mathcal{I}$. Then, it is clear that $\zeta(\lambda)$ is finite, for all $0 < \lambda \le 1$. Because of the way the constant $\mathcal{H}$ enters into the definition, it remains to show that $\zeta(1)$ admits an absolute upper bound, independent of how $\mathcal{I}$ is chosen.

It is the consequence of Lemma 4.6 that there holds

$$\zeta(\lambda) \le C\lambda + (1+\sqrt{2})\zeta(\lambda/4) , \qquad 0 < \lambda < 1 .$$

Iterating this inequality beginning at $\lambda = 1$ gives us

$$\zeta(1) \le C + (1+\sqrt{2})\zeta(1/4) \le \cdots \le C\sum_{t=0}^{\infty}\Big[\tfrac{1+\sqrt{2}}{4}\Big]^t \le 4C\,.$$

So we have established an absolute upper bound on $\zeta(1)$. □

4.3. **Proof of Lemma 4.6.** We restate the conclusion of Lemma 4.6 to more closely follow the line of argument to follow. The collection $\mathcal{Q}$ can be partitioned into two collections $\mathcal{Q}^{\text{large}}$ and $\mathcal{Q}^{\text{small}}$ such that

(1) $\mathbf{B}_{\mathcal{Q}^{\text{large}}} \lesssim \tau$, where $\tau = \text{size}(\mathcal{Q})$.
(2) $\mathcal{Q}^{\text{small}} = \mathcal{Q}_1^{\text{small}} \cup \mathcal{Q}_2^{\text{small}}$.
(3) The collection $\mathcal{Q}_1^{\text{small}}$ is admissible, and $\text{size}(\mathcal{Q}_1^{\text{small}}) \le \frac{\tau}{4}$.
(4) For a collection of dyadic intervals $\mathcal{L}$, the collection $\mathcal{Q}_2^{\text{small}}$ is the union of mutually orthogonal admissible collections $\mathcal{Q}_{2,L}^{\text{small}}$, for $L \in \mathcal{L}$, with

$$\text{size}(\mathcal{Q}_{2,L}^{\text{small}}) \le \tfrac{\tau}{4}\,, \qquad L \in \mathcal{L}\,.$$

Thus, we have by inequality (4.5) for mutually orthogonal collections,

$$\begin{aligned}\mathbf{B}_{\mathcal{Q}} &\le \mathbf{B}_{\mathcal{Q}^{\text{large}}} + \mathbf{B}_{\mathcal{Q}_1^{\text{small}} \cup \mathcal{Q}_2^{\text{small}}}\\ &\le \mathbf{B}_{\mathcal{Q}^{\text{large}}} + \mathbf{B}_{\mathcal{Q}_1^{\text{small}}} + \mathbf{B}_{\mathcal{Q}_2^{\text{small}}}\\ &\le C\tau + (1+\sqrt{2})\max\Big\{\mathbf{B}_{\mathcal{Q}_1^{\text{small}}}, \sup_{L\in\mathcal{L}} \mathbf{B}_{\mathcal{Q}_{2,L}^{\text{small}}}\Big\}\,.\end{aligned}$$

This, with the properties of size listed above prove Lemma 4.6 as stated, after a trivial re-indexing.

All else flows from this construction of a subset $\mathcal{L}$ of dyadic subintervals of I_0. The initial intervals in $\mathcal{L}$ are the minimal intervals $K \in \tilde{\mathcal{Q}}_1 \cup \mathcal{Q}_2$ such that

$$\frac{P(\sigma(I_0 - K), K)^2}{|K|^2} \sum_{J\in\mathcal{Q}_2\,:\,J\subset K} \langle x, h_J^w\rangle_w^2 \ge \frac{\tau^2}{16}\sigma(K)\,. \tag{4.8}$$

Since $\text{size}(\mathcal{Q}) = \tau$, there are such intervals K.

Initialize $\mathcal{S}$ (for 'stock' or 'supply') to be all the dyadic intervals in $\tilde{\mathcal{Q}}_1 \cup \mathcal{Q}_2$ which are not contained in any element of $\mathcal{L}$. In the recursive step, let $\mathcal{L}'$ be the minimal elements $S \in \mathcal{S}$ such that

$$\sum_{J\in\mathcal{Q}_2\,:\,J\subset S} \langle x, h_J^w\rangle_w^2 \ge \rho \sum_{\substack{L\in\mathcal{L}\,:\,L\subset S\\ L \text{ is maximal}}} \sum_{J\in\mathcal{Q}_2\,:\,J\subset L} \langle x, h_J^w\rangle_w^2\,, \qquad \rho = \tfrac{17}{16}\,. \tag{4.9}$$

(The inequality would be trivial if $\rho = 1$.) If $\mathcal{L}'$ is empty the recursion stops. Otherwise, update $\mathcal{L} \leftarrow \mathcal{L} \cup \mathcal{L}'$, and $\mathcal{S} \leftarrow \{K \in \mathcal{S} \,:\, K \not\subset L\ \forall L \in \mathcal{L}\}$.

Once the recursion stops, report the collection $\mathcal{L}$. It has this crucial property: For $L \in \mathcal{L}$, and integers $t \ge 1$,

$$\sum_{L'\,:\,\pi_{\mathcal{L}}^t L' = L} \sum_{J\in\mathcal{Q}_2\,:\,J\subset L'} \langle x, h_J^w\rangle_w^2 \le \rho^{-t} \sum_{J\in\mathcal{Q}_2\,:\,J\subset L} \langle x, h_J^w\rangle_w^2\,. \tag{4.10}$$

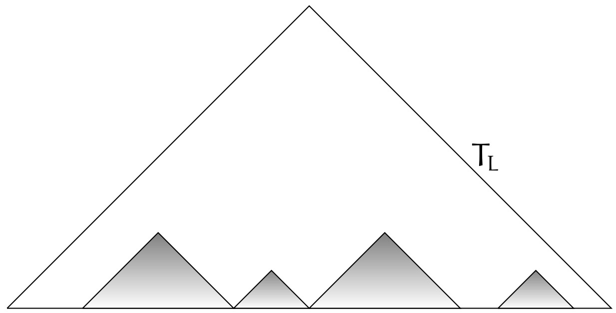

FIGURE 1. The shaded smaller tents have been selected, and T_L is the minimal tent with $\mu(T_L)$ larger than ρ times the μ-measure of the shaded tents.

Indeed, in the case of $t = 1$, is the selection criteria for membership in $\mathcal{L}$, and a simple induction proves the statement for all $t \geq 1$.

4.11. *Remark.* The selection of $\mathcal{L}$ can be understood as a familiar argument concerning Carleson measures, although there is no such object in this argument. Consider the measure μ on $\mathbb{R}^2_+$ given as a sum of point masses given by

$$\mu := \sum_{J \in \mathcal{Q}_2 \,:\, J \subset I_0} \langle x, h^w_J \rangle^2_w \delta_{(x_J, |J|)}\,, \qquad x_J \text{ is the center of } J.$$

The tent over L is the triangular region $T_L := \{(x, y) \,:\, |x - x_L| \leq |L| - y\}$, so that

$$\mu(T_L) = \sum_{J \in \mathcal{Q}_2 \,:\, J \subset L} \langle x, h^w_J \rangle^2_w\,.$$

Then, the selection rule for membership in $\mathcal{L}$ can be understood as taking the minimal tent T_L such that $\mu(T_L)$ is bigger than ρ times the μ-measure of the selected tents. See Figure 1.

The decomposition of $\mathcal{Q}$ is based upon the relation of the pairs to the collection $\mathcal{L}$, namely a pair $\tilde{Q}_1, Q_2$ can (a) both have the same parent in $\mathcal{L}$; (b) have distinct parents in $\mathcal{L}$; (c) Q_2 can have a parent in $\mathcal{L}$, but not $\tilde{Q}_1$; and (d) Q_2 does not have a parent in $\mathcal{L}$.

A particularly vexing aspect of the stopping form is the linkage between the martingale difference on g, which is given by J, and the argument of the Hilbert transform, $I_0 - I_J$. The 'large' collections constructed below will, in a certain way, decouple the J and the $I_0 - I_J$, enough so that norm of the associated bilinear form can be estimated by the size of $\mathcal{Q}$.

In the 'small' collections, there is however no decoupling, but critically, both the size of the collections is smaller, and that the estimate is given in terms of the supremum in (4.7).

Pairs comparable to $\mathcal{L}$. Define

$$\mathcal{Q}_{L,t} := \{Q \in \mathcal{Q} \,:\, \pi_{\mathcal{L}} \tilde{Q}_1 = \pi^t_{\mathcal{L}} Q_2 = L\}, \qquad L \in \mathcal{L}\,,\ t \in \mathbb{N}\,.$$

These are admissible collections, as the convexity property in Q_1, holding Q_2 constant, is clearly inherited from $\mathcal{Q}$. Now, observe that for each $t \in \mathbb{N}$, the collections $\{\mathcal{Q}_{L,t} \,:\, L \in \mathcal{L}\}$ are mutually orthogonal. The collection of intervals $(\mathcal{Q}_{L,t})_2$ are obviously disjoint in $L \in \mathcal{L}$, with $t \in \mathbb{N}$ held

fixed. And, since membership in these collections is determined in the first coordinate by the interval $\tilde{Q}_1$, and the two children of Q_1 can have two different parents in $\mathcal{L}$, a given interval I can appear in at most two collections $(\widetilde{\mathcal{Q}_{L,t}})_1$, as $L \in \mathcal{L}$ varies, and $t \in \mathbb{N}$ held fixed.

Define $\mathcal{Q}_1^{\text{small}}$ to be the union over $L \in \mathcal{L}$ of the collections

$$\mathcal{Q}_{1,L}^{\text{small}} := \{Q \in \mathcal{Q}_{L,1} \;:\; \tilde{Q}_1 \neq L\}\,.$$

Note in particular that we have only allowed $t = 1$ above, and $\tilde{Q}_1 = L$ is not allowed. For these collections, we need only verify that

$$\text{size}(\mathcal{Q}_{1,L}^{\text{small}}) \leq \sqrt{(\rho - 1)} \cdot \tau = \frac{\tau}{4}\,, \qquad L \in \mathcal{L}\,,\ t \in \mathbb{N}\,. \tag{4.12}$$

Proof. An interval $K \in (\widetilde{\mathcal{Q}_{1,L}^{\text{small}}})_1 \cup \mathcal{Q}_2$ is not in $\mathcal{L}$, by construction. Suppose that K does not contain any interval in $\mathcal{L}$. By the selection of the initial intervals in $\mathcal{L}$, the minimal intervals in $\tilde{\mathcal{Q}}_1 \cup \mathcal{Q}_2$ which satisfy (4.8), it follows that the interval K must fail (4.8). And so we are done.

Thus, K contains some element of $\mathcal{L}$, whence the inequality (4.9) must fail. Namely, rearranging that inequality,

$$\sum_{\substack{J \in \mathcal{Q}_2 \,:\, \pi_{\mathcal{L}} J = L \\ J \subset K}} \langle x, h_J^w \rangle_w^2 \leq (\rho - 1) \sum_{\substack{L' \in \mathcal{L} \,:\, L' \subset K \\ L' \text{ is maximal}}} \sum_{J \in \mathcal{Q}_2 \,:\, J \subset L} \langle x, h_J^w \rangle_w^2\,.$$

Recall that $\rho - 1 = \frac{1}{16}$. We can estimate

$$\begin{aligned} \sum_{\substack{J \in \mathcal{Q}_2 \,:\, \pi_{\mathcal{L}} J = L \\ J \subset K}} \langle x, h_J^w \rangle_w^2 &\leq \frac{1}{16} \sum_{J \in \mathcal{Q}_2 \,:\, J \subset L} \langle x, h_J^w \rangle_w^2 \\ &\leq \frac{\tau^2}{16} \cdot \frac{|K|^2 \cdot \sigma(K)}{P(\sigma(L - K), K)^2}\,. \end{aligned}$$

The last inequality follows from the definition of size, and finishes the proof of (4.12). ◻

The collections below are the first contribution to $\mathcal{Q}^{\text{large}}$. Take $\mathcal{Q}_1^{\text{large}} := \cup\{\mathcal{Q}_{1,L}^{\text{large}} \;:\; L \in \mathcal{L}\}$, where

$$\mathcal{Q}_{1,L}^{\text{large}} := \{Q \in \mathcal{Q}_{L,1} \;:\; \tilde{Q}_1 = L\}\,.$$

Note that Lemma 4.17 applies to this Lemma, take the collection $\mathcal{S}$ of that Lemma to be $\{L\}$, and the quantity η in (4.18) satisfies $\eta \lesssim \tau = \text{size}(\mathcal{Q})$, by inspection. From the mutual orthogonality (4.5), we then have

$$\mathbf{B}_{\mathcal{Q}_1^{\text{large}}} \leq \sqrt{2} \sup_{L \in \mathcal{L}} \mathbf{B}_{\mathcal{Q}_{1,L}^{\text{large}}} \lesssim \tau\,.$$

The collections $\mathcal{Q}_{L,t}$, for $L \in \mathcal{L}$, and $t \geq 2$ are the second contribution to $\mathcal{Q}^{\text{large}}$, namely

$$\mathcal{Q}_2^{\text{large}} := \bigcup_{L \in \mathcal{L}} \bigcup_{t \geq 2} \mathcal{Q}_{L,t}\,.$$

For them, we need to estimate $\mathbf{B}_{\mathcal{Q}_{L,t}}$.

$$\mathbf{B}_{\mathcal{Q}_{L,t}} \lesssim \rho^{-t/2}\tau\,. \tag{4.13}$$

From this, we can conclude from (4.5) that

$$\begin{aligned}\mathbf{B}_{\mathcal{Q}_2^{\text{large}}} &\le \sum_{t\ge 2} \mathbf{B}_{\bigcup\{\mathcal{Q}_{L,t}\,:\, L\in\mathcal{L}\}} \\ &\le \sqrt{2}\sum_{t\ge 2}\sup_{L\in\mathcal{L}} \mathbf{B}_{\mathcal{Q}_{L,t}} \lesssim \tau\sum_{t\ge 2}\rho^{-t/2} \lesssim \tau\,.\end{aligned}$$

Proof of (4.13). For $L \in \mathcal{L}$, let $\mathcal{S}_L$, the $\mathcal{L}$-children of L. For each $Q \in \mathcal{Q}_{L,t}$, we must have $Q_2 \subset \pi_{\mathcal{S}_L} Q_2 \subset \tilde{Q}_1$. Then, divide the collection $\mathcal{Q}_{L,t}$ into three collections $\mathcal{Q}^{\ell}_{L,t}$, $\ell = 1,2,3$, where

$$\begin{aligned}\mathcal{Q}^1_{L,t} &:= \{Q \in \mathcal{Q}_{L,t} \;:\; Q_2 \Subset \pi_{\mathcal{S}_L} Q_2\}\,,\\ \mathcal{Q}^2_{L,t} &:= \{Q \in \mathcal{Q}_{L,t} \;:\; Q_2 \not\Subset \pi_{\mathcal{S}_L} Q_2 \Subset \tilde{Q}_1\}\,,\end{aligned}$$

and $\mathcal{Q}^3_{L,t} := \mathcal{Q}_{L,t} - (\mathcal{Q}^1_{L,t} \cup \mathcal{Q}^2_{L,t})$ is the complementary collection. Notice that $\mathcal{Q}^1_{L,t}$ equals the whole collection $\mathcal{Q}_{L,t}$ for $t > r+1$.

We treat them in turn. The collections $\mathcal{Q}^1_{L,t}$ fit the hypotheses of Lemma 4.17, just take the collection of intervals $\mathcal{S}$ of that Lemma to be $\mathcal{S}_L$. It follows that $\mathbf{B}_{\mathcal{Q}^1_{L,t}} \lesssim \boldsymbol{\beta}(t)$, where the latter is the best constant in the inequality

$$\sum_{J\in(\mathcal{Q}_{L,t})_2\,:\,J\Subset K} P(\sigma(I_0-K),J)^2\Big\langle \frac{x}{|J|}, h^w_J\Big\rangle_w^2 \le \boldsymbol{\beta}(t)^2\sigma(K)\,, \qquad K\in\mathcal{S}_L\,,\ L\in\mathcal{L}\,,\ t\ge 2\,. \tag{4.14}$$

There is an observation about the Poisson integral terms that we need. For K as above, and $J \subset L' \Subset K$, note that by goodness of L',

$$\operatorname{dist}(J, I_0 - K) \ge \operatorname{dist}(L', I_0-K) > |L'|^{\epsilon}|K|^{1-\epsilon} \ge 2^{(r+1)(1-\epsilon)}|L'|\,.$$

From the definition of the Poisson integral, one sees that

$$\frac{P(\sigma(I_0-K),J)}{|J|} \lesssim \frac{P(\sigma(I_0-K),L')}{|L'|}\,. \tag{4.15}$$

We have the estimate without decay in t, $\boldsymbol{\beta}(t) \lesssim \operatorname{size}(\mathcal{Q})$. Indeed, for K as in (4.14), let $\mathcal{J}^*$ be the maximal intervals with $J^* \in (\mathcal{Q}_{L,t})_2$ and $J^* \Subset K$. Now, $\mathcal{J}^*$ is contained in the collection of intervals over which we test the size of $\mathcal{Q}$, hence by (4.15),

$$\begin{aligned}\text{LHS}(4.14) &= \sum_{J^*\in\mathcal{J}^*}\ \sum_{J\in(\mathcal{Q}_{L,t})_2\,:\,J\subset J^*} P(\sigma(I_0-K),J)^2\Big\langle\frac{x}{|J|}, h^w_J\Big\rangle_w^2\\ &\lesssim \sum_{J^*\in\mathcal{J}^*}\frac{P(\sigma(I_0-K),J^*)^2}{|J^*|^2}\sum_{J\in(\mathcal{Q}_{L,t})_2\,:\,J\subset J^*}\langle x, h^w_J\rangle_w^2\\ &\lesssim \tau^2\sum_{J^*\in\mathcal{J}^*}\sigma(J^*) \lesssim \tau^2\sigma(K)\,.\end{aligned}$$

This proves the claim, and we use the estimate for $t \le r+3$, say. (Recall that r is a fixed integer.)

In the case of $t > r+3$, the essential property is (4.10). The left hand side of (4.14) is dominated by the sum below. Note that we index the sum first over L', which are $r+1$-fold $\mathcal{L}$-children of K, whence $L' \Subset K$, followed by $t-r-2$-fold $\mathcal{L}$-children of L'.

$$\begin{aligned}
\sum_{\substack{L'\in\mathcal{L}\\ \pi_{\mathcal{L}}^{r+1}L'=K}} \sum_{\substack{L''\in\mathcal{L}\\ \pi_{\mathcal{L}}^{t-r-2}L''=L'}} \sum_{J\in\mathcal{Q}_2\,:\,J\subset L''} & P(\sigma(I_0-K),J)^2\Big\langle \frac{x}{|J|}, h_J^w\Big\rangle_w^2 \\
&\overset{(4.15)}{\le} \sum_{\substack{L'\in\mathcal{L}\\ \pi_{\mathcal{L}}^{r+1}L'=K}} \frac{P(\sigma(I_0-K),L')^2}{|L'|^2} \sum_{\substack{L''\in\mathcal{L}\\ \pi_{\mathcal{L}}^{t-r-2}L''=L'}} \langle x, h_J^w\rangle_w^2 \\
&\overset{(4.10)}{\lesssim} \rho^{-t+r+2} \sum_{\substack{L'\in\mathcal{L}\\ \pi_{\mathcal{L}}^{r+1}L'=K}} \frac{P(\sigma(I_0-K),L')^2}{|L'|^2} \sum_{J\in\mathcal{Q}_2\,:\,J\subset L'} \langle x, h_J^w\rangle_w^2 \\
&\lesssim \rho^{-t}\tau^2 \sum_{\substack{L'\in\mathcal{L}\\ \pi_{\mathcal{L}}^{r+1}L'=K}} \sigma(L') \lesssim \tau^2\rho^{-t}\sigma(K)\,.
\end{aligned}$$

We have also used (4.15), and then the central property (4.10) following from the construction of $\mathcal{L}$, finally appealing to the definition of size. Hence, $\boldsymbol{\beta}(t) \lesssim \tau^2\rho^{-t}$. This completes the analysis of $\mathcal{Q}^1_{L,t}$.

We need only consider the collections $\mathcal{Q}^2_{L,t}$ for $1 \le t \le r+1$, and they fall under the scope of Lemma 4.22. And, we see immediately that we have $\mathbf{B}_{\mathcal{Q}^2_{L,t}} \lesssim \tau$. Similarly, we need only consider the collections $\mathcal{Q}^3_{L,t}$ for $1 \le t \le r+1$. It follows that we must have $2^r \le |Q_1|/|Q_2| \le 2^{2r+2}$. Namely, this ratio can take only one of a finite number of values, implying that Lemma 4.24 applies easily to this case to complete the proof. □

Pairs not strictly comparable to $\mathcal{L}$. It remains to consider the pairs $Q \in \mathcal{Q}$ such that $\tilde{Q}_1$ does not have a parent in $\mathcal{L}$. The collection $\mathcal{Q}_2^{\mathrm{small}}$ is taken to be the (much smaller) collection

$$\mathcal{Q}_2^{\mathrm{small}} := \{Q \in \mathcal{Q} \,:\, Q_2 \text{ does not have a parent in } \mathcal{L}\}\,.$$

Observe that $\mathrm{size}(\mathcal{Q}_2^{\mathrm{small}}) \le \sqrt{(\rho-1)}\tau \le \frac{\tau}{4}$. This is as required for this collection.[7]

Proof. Suppose $\eta < \mathrm{size}(\mathcal{Q}_2^{\mathrm{small}})$. Then, there is an interval $K \in (\widetilde{\mathcal{Q}_1^{\mathrm{small}}})_1 \cup (\mathcal{Q}_2^{\mathrm{small}})_2$ so that

$$\eta^2\sigma(K) \le \frac{P(\sigma(I_0-K),K)^2}{|K|^2} \sum_{\substack{J\in(\mathcal{Q}_2^{\mathrm{small}})_2\\ J\subset K}} \langle x, h_J^w\rangle_w^2\,.$$

Suppose that K does not contain any interval in $\mathcal{L}$. It follows from the initial intervals added to $\mathcal{L}$, see (4.8), that we must have $\eta \le \frac{\tau}{4}$.

[7] The collections $\mathcal{Q}_1^{\mathrm{small}}$ and $\mathcal{Q}_2^{\mathrm{small}}$ are also mutually orthogonal, but this fact is not needed for our proof.

Thus, K contains an interval in $\mathcal{L}$. This means that K must fail the inequality (4.9). Therefore, we have

$$\eta^2\sigma(K) \le (\rho-1)\frac{P(\sigma(I_0-K),K)^2}{|K|^2}\sum_{\substack{J\in\mathcal{Q}_2\\ J\subset K}}\langle x, h_J^w\rangle_w^2 \le \frac{\tau^2}{16}\sigma(K)\,.$$

This relies upon the definition of size, and proves our claim. □

For the pairs not yet in one of our collections, it must be that Q_2 has a parent in $\mathcal{L}$, but not $\tilde{Q}_1$. Using $\mathcal{L}^*$, the maximal intervals in $\mathcal{L}$, divide them into the three collections

$$\mathcal{Q}_3^{\text{large}} := \{Q\in\mathcal{Q} \,:\, Q_2 \Subset \pi_{\mathcal{L}^*}Q_2 \subset \tilde{Q}_1\},$$
$$\mathcal{Q}_4^{\text{large}} := \{Q\in\mathcal{Q} \,:\, Q_2 \not\Subset \pi_{\mathcal{L}^*}Q_2 \Subset \tilde{Q}_1\},$$
$$\mathcal{Q}_5^{\text{large}} := \{Q\in\mathcal{Q} \,:\, Q_2 \not\Subset \pi_{\mathcal{L}^*}Q_2 \subsetneq \tilde{Q}_1\,, \text{and } \pi_{\mathcal{L}^*}Q_2 \not\Subset \tilde{Q}_1\}.$$

Observe that Lemma 4.17 applies to give

$$\mathbf{B}_{\mathcal{Q}_3^{\text{large}}} \lesssim \tau\,. \tag{4.16}$$

Take the collection $\mathcal{S}$ of Lemma 4.17 to be $\mathcal{L}^*$, and note that the bound in that Lemma is given by η, as defined in (4.18), which by construction is less than $\tau = \text{size}(\mathcal{Q})$.

Observe that Lemma 4.22 applies to show that the estimate (4.16) holds for $\mathcal{Q}_4^{\text{large}}$. Take $\mathcal{S}$ of that Lemma to be $\mathcal{L}^*$. The estimate from Lemma 4.22 is given in terms of η, as defined in (4.23). But, is at most τ.

In the last collection, $\mathcal{Q}_5^{\text{large}}$, notice that the conditions placed upon the pair implies that $|Q_1| \le 2^{2r+2}|Q_2|$, for all $Q\in\mathcal{Q}_5^{\text{large}}$. It therefore follows from a straight forward application of Lemma 4.24, that (4.16) holds for this collection as well.

4.4. **Upper Bounds on the Stopping Form.** We have three lemmas that prove upper bounds on the norm of the stopping form in situations in which there is a measure of decoupling between the martingale difference on g, and the argument of the Hilbert transform.

4.17. **Lemma.** *Let $\mathcal{S}$ be a collection of pairwise disjoint intervals in I_0. Let $\mathcal{Q}$ be admissible such that for each $Q\in\mathcal{Q}$, there is an $S\in\mathcal{S}$ with $Q_2\Subset S\subset\tilde{Q}_1$. Then, there holds*

$$|B_{\mathcal{Q}}(f,g)| \lesssim \eta\|f\|_\sigma\|g\|_w\,,$$

$$\text{where}\quad \eta^2 := \sup_{S\in\mathcal{S}}\frac{1}{\sigma(S)}\sum_{J\in\mathcal{Q}_2\,:\,J\Subset S} P(\sigma(I_0-S),J)^2\Big\langle\frac{x}{|J|},h_J^w\Big\rangle_w^2\,. \tag{4.18}$$

(Note that size$(\mathcal{Q})$ *need not control* η.*)*

Proof. An interesting part of the proof is that it depends very much on cancellative properties of the martingale differences of f. (Absolute values must be taken *outside* the sum defining the stopping form!)

Assume that the Haar support of f is contained in $\mathcal{Q}_1$. Take $\mathcal{F}$ and $\alpha_f(\cdot)$ to be stopping data defined in this way. First, add to $\mathcal{F}$ the interval I_0, and set $\alpha_f(I_0) := \mathbb{E}_{I_0}^\sigma|f|$. Inductively, if $F\in\mathcal{F}$

is minimal, add to $\mathcal{F}$ the maximal children F' such that $\alpha_f(F') := \mathbb{E}^{\sigma}_{F'}|f| > 4\alpha_f(F)$. Note that the inequality (3.9) holds for this choice of $\mathcal{F}$ and α_f, so that the quasi-orthogonality argument (3.10) is available to us.

Write the bilinear form as

$$B_{\mathcal{Q}}(f,g) = \sum_{J} \langle H_\sigma \varphi_J, \Delta^w_J g\rangle_w$$

$$\text{where} \quad \varphi_J := \sum_{Q\in\mathcal{Q}\,:\,Q_2=J} \mathbb{E}^{\sigma}_J \Delta^{\sigma}_{Q_1} f \cdot (I_0 - \tilde{Q}_1)\,. \tag{4.19}$$

The function φ_J is well-behaved. For any $J \in \mathcal{Q}_2$, $|\varphi_J| \lesssim \alpha_f(\pi_{\mathcal{F}} J)\Delta J$. In this definition, $\Delta J := \bigcup\{I_0 - \tilde{Q}_1 \ :\ Q \in \mathcal{Q}\,,\ Q_2 = J\}$. Indeed, at each point $x \in \Delta J$, the sum defining $\varphi_J(x)$ is over pairs Q such that $Q_2 = J$ and $x \in I_0 - \tilde{Q}_1$. By the convexity property of admissible collections, the sum is over consecutive martingale differences of f. The basic telescoping property of these differences shows that the sum is bounded by the stopping value $\alpha_f(\pi_{\mathcal{F}} J)$. Let I^* be the maximal interval of the form $\tilde{Q}_1$ with $x \in I_0 - \tilde{Q}_1$, and let I_* be the child of the minimal such interval which contains J. Then,

$$\begin{aligned} |\varphi_J(x)| &= \Big| \sum_{\substack{Q\in\mathcal{Q}\,:\,Q_2=J\\ x\in I-\tilde{Q}_1}} \mathbb{E}^{\sigma}_J \Delta^{\sigma}_{Q_1} f(x)\Big| \\ &= \big|\mathbb{E}^{\sigma}_{I^*} f - \mathbb{E}^{\sigma}_{I_*} f\big| \lesssim \alpha_f(\pi_{\mathcal{F}} J)(I_0 - S)\,, \end{aligned} \tag{4.20}$$

where S is the $\mathcal{S}$-parent of J.

We can estimate as below, for $F \in \mathcal{F}$:

$$\begin{aligned} \Xi(F) &:= \Big| \sum_{Q\in\mathcal{Q}\,:\,\pi_{\mathcal{F}} Q_2=F} \mathbb{E}_{Q_2} \Delta^{\sigma}_{Q_1} f \cdot \langle H_\sigma (I_0 - \tilde{Q}_1), \Delta^w_J g\rangle_w \Big| \\ &\overset{(4.19)}{=} \Big| \sum_{J\in\mathcal{Q}_2\,:\,\pi_{\mathcal{F}} J=F} \langle H_\sigma \varphi_J, \Delta^w_J g\rangle_w \Big| \\ &\overset{(4.20)}{\lesssim} \alpha_f(F) \sum_{\substack{S\in\mathcal{S}\\ \pi_{\mathcal{F}} S=F}} \sum_{\substack{J\in\mathcal{Q}_2\\ J\subset S}} P(\sigma(I_0 - S), J) \Big| \Big\langle \frac{x}{|J|}, \Delta^w_J g \Big\rangle_w \Big| \\ &\lesssim \alpha_f(F) \Big[\sum_{\substack{S\in\mathcal{S}\\ \pi_{\mathcal{F}} S=F}} \sum_{\substack{J\in\mathcal{Q}_2\\ J\subset S}} P(\sigma(I_0 - S), J)^2 \Big\langle \frac{x}{|J|}, h^w_J \Big\rangle_w^2 \times \sum_{\substack{J\in\mathcal{Q}_2\\ \pi_{\mathcal{F}} J=F}} \hat{g}(J)^2 \Big]^{1/2} \\ &\overset{(4.18)}{\lesssim} \eta\alpha_f(F) \Big[\sum_{\substack{S\in\mathcal{S}\\ \pi_{\mathcal{F}} S=F}} \sigma(S) \times \sum_{\substack{J\in\mathcal{Q}_2\\ \pi_{\mathcal{F}} J=F}} \hat{g}(J)^2 \Big]^{1/2} \\ &\lesssim \eta\alpha_f(F)\sigma(F)^{1/2} \Big[\sum_{J\in\mathcal{Q}_2\,:\,\pi_{\mathcal{F}} J=F} \hat{g}(J)^2 \Big]^{1/2}. \end{aligned}$$

The top line follows from (4.19). In the second, we appeal to (4.20) and monotonicity (2.3), the latter being available to us since $J \subset S$ implies $J \Subset S$, by hypothesis. We also take advantage of the strong assumptions on the intervals in $\mathcal{Q}_2$: If $J \in \mathcal{Q}_2$, we must have $\pi_{\mathcal{F}} J = \pi_{\mathcal{F}}(\pi_{\mathcal{S}} J)$. The third line is Cauchy–Schwarz, followed by the appeal to the hypothesis (4.18), while the last line uses the fact that the intervals in $\mathcal{S}$ are pairwise disjoint.

The quasi-orthogonality argument (3.10) completes the proof, namely we have

$$\sum_{F \in \mathcal{F}} \Xi(F) \lesssim \eta \|f\|_\sigma \|g\|_w . \tag{4.21}$$

□

4.22. **Lemma.** *Let $\mathcal{S}$ be a collection of pairwise disjoint intervals in I_0. Let $\mathcal{Q}$ be admissible such that for each $Q \in \mathcal{Q}$, there is an $S \in \mathcal{S}$ with $Q_2 \subset S \Subset \tilde{Q}_1$. Then, there holds*

$$|B_{\mathcal{Q}}(f, g)| \lesssim \eta \|f\|_\sigma \|g\|_w ,$$

$$\text{where} \quad \eta^2 := \sup_{S \in \mathcal{S}} \frac{P(\sigma(Q_1 - \pi_{\tilde{\mathcal{Q}}_1} S), S)^2}{\sigma(S)|S|^2} \sum_{J \in \mathcal{Q}_2 : J \subset S} \langle x, h_J^w \rangle_w^2 . \tag{4.23}$$

Proof. Construct stopping data $\mathcal{F}$ and $\alpha_f(\cdot)$ as in the proof of Lemma 4.17. The fundamental inequality (4.20) is again used. Then, by the monotonicity principle (2.3), there holds for $F \in \mathcal{F}$,

$$\begin{aligned}
\Xi(F) :=& \Bigl| \sum_{Q \in \mathcal{Q} : \pi_{\mathcal{F}} Q_2 = F} \mathbb{E}_{Q_2} \Delta^\sigma_{Q_1} f \cdot \langle H_\sigma(I_0 - \tilde{Q}_1), \Delta^w_{Q_2} g \rangle_w \Bigr| \\
&\lesssim \alpha_f(F) \sum_{S \in \mathcal{S} : \pi_{\mathcal{F}} S = F} P(\sigma(I_0 - \pi_{\tilde{\mathcal{Q}}_1} S), S) \sum_{J \in \mathcal{Q}_2 : J \subset S} \Bigl\langle \frac{x}{|S|}, h_J^w \Bigr\rangle_w \cdot |\hat{g}(J)| \\
&\lesssim \alpha_f(F) \Bigl[\sum_{S \in \mathcal{S} : \pi_{\mathcal{F}} S = F} P(\sigma(I_0 - \pi_{\tilde{\mathcal{Q}}_1} S), S)^2 \sum_{J \in \mathcal{Q}_2 : J \subset S} \Bigl\langle \frac{x}{|S|}, h_J^w \Bigr\rangle_w^2 \times \sum_{J \in \mathcal{Q}_2 : J \subset S} \hat{g}(J)^2 \Bigr]^{1/2} \\
&\lesssim \eta \alpha_f(F) \Bigl[\sum_{S \in \mathcal{S} : \pi_{\mathcal{F}} S = F} \sigma(S) \times \sum_{J \in \mathcal{Q}_2 : J \subset S} \hat{g}(J)^2 \Bigr]^{1/2} \\
&\lesssim \eta \alpha_f(F) \sigma(F)^{1/2} \Bigl[\sum_{J \in \mathcal{Q}_2 : \pi_{\mathcal{F}} J = F} \hat{g}(J)^2 \Bigr]^{1/2} .
\end{aligned}$$

After the monotonicity principle (2.3), we have used Cauchy–Schwarz, and the definition of η. The quasi-orthogonality argument (3.9) then completes the analysis of this term, see (4.21). □

The last Lemma that we need is elementary, and is contained in the methods of [18].

4.24. **Lemma.** *Let $u \geq r + 1$ be an integer, and $\mathcal{Q}$ be an admissible collection of pairs such that $|Q_1| = 2^u |Q_2|$ for all $Q \in \mathcal{Q}$. There holds*

$$|B_{\mathcal{Q}}(f, g)| \lesssim \operatorname{size}(\mathcal{Q}) \|f\|_\sigma \|g\|_w .$$

Proof. Recall the form of the stopping form in (4.1). Observe, from inspection of the definition of the Haar function (2.1), that

$$|\mathbb{E}^\sigma_{I_J} \Delta^\sigma_I f| \leq \frac{|\hat{f}(I)|}{\sigma(I_J)^{1/2}} .$$

Then, we have, keeping in mind that I_J is one or the other of the two children of I,

$$\begin{aligned}
|B_{\mathcal{Q}}(f,g)| &\leq \sum_{I \in \mathcal{Q}_1} |\hat{f}(I)| \sum_{J \,:\, (I,J) \in \mathcal{Q}} \sigma(I_J)^{-1/2} P(\sigma(I_0 - I_J), J) \Big\langle \frac{x}{|J|}, h^w_J \Big\rangle_w |\hat{g}(J)| \\
&\leq \|f\|_\sigma \Bigg[\sum_{I \in \mathcal{Q}_1} \Bigg[\sum_{J \,:\, (I,J) \in \mathcal{Q}} \frac{1}{\sigma(I_J)} P(\sigma(I_0 - I_J), J) \Big\langle \frac{x}{|J|}, h^w_J \Big\rangle_w |\hat{g}(J)| \Bigg]^2 \Bigg]^{1/2} \\
&\leq \mathsf{size}(\mathcal{Q}) \|f\|_\sigma \|g\|_w
\end{aligned}$$

This follows immediately from Cauchy–Schwarz, and the fact that for each $J \in \mathcal{Q}_2$, there is a unique $I \in \mathcal{Q}_1$ such that the pair (I, J) contribute to the sum above. ◻

References

[1] Yurii Belov, Tesfa Y. Mengestie, and Kristian Seip, *Discrete Hilbert transforms on sparse sequences*, Proc. Lond. Math. Soc. (3) **103** (2011), no. 1, 73–105.

[2] Lennart Carleson, *On convergence and growth of partial sums of Fourier series*, Acta Math. **116** (1966), 135–157.

[3] M. Cotlar and C. Sadosky, *On some L^p versions of the Helson-Szegő theorem*, (Chicago, Ill., 1981), Wadsworth Math. Ser., Wadsworth, Belmont, CA, 1983, pp. 306–317.

[4] David Cruz-Uribe, *The invertibility of the product of unbounded Toeplitz operators*, Integral Equations Operator Theory **20** (1994), no. 2, 231–237.

[5] Guy David and Jean-Lin Journé, *A boundedness criterion for generalized Calderón-Zygmund operators*, Ann. of Math. (2) **120** (1984), no. 2, 371–397.

[6] Charles Fefferman, *Pointwise convergence of Fourier series*, Ann. of Math. (2) **98** (1973), 551–571.

[7] Tuomas P. Hytönen, *The sharp weighted bound for general Calderón-Zygmund operators*, Ann. of Math. (2) **175** (2012), no. 3, 1473–1506.

[8] Michael T. Lacey, Eric T. Sawyer, and Ignacio Uriarte-Tuero, *A Two Weight Inequality for the Hilbert transform Assuming an Energy Hypothesis*, J Funct Anal **263** (2012), 305–363.

[9] Michael Lacey, Eric T. Sawyer, and Ignacio Uriarte-Tuero, *A characterization of two weight norm inequalities for maximal singular integrals with one doubling measure*, Anal. PDE **5** (2012), no. 1, 1–60.

[10] Michael T. Lacey, Eric T. Sawyer, Ignacio Uriarte-Tuero, and Chun-Yen Shen, *The Two Weight Inequality for Hilbert Transform, Coronas, and Energy Conditions*, available at http://www.arxiv.org/abs/1108.2319.

[11] ———, *Two Weight Inequality for the Hilbert Transform: A Real Variable Characterization*, Submitted, available at http://www.arxiv.org/abs/1201.4319.

[12] Michael T. Lacey and Christoph Thiele, *A proof of boundedness of the Carleson operator*, Math. Res. Lett. **7** (2000), 361–370.

[13] Benjamin Muckenhoupt and Richard L. Wheeden, *Two weight function norm inequalities for the Hardy-Littlewood maximal function and the Hilbert transform*, Studia Math. **55** (1976), no. 3, 279–294.

[14] F. Nazarov, *A counterexample to Sarason's conjecture*, Preprint, MSU (1997), available at http://www.math.msu.edu/~fedja/prepr.html.

[15] F. Nazarov, S. Treil, and A. Volberg, *Cauchy integral and Calderón-Zygmund operators on nonhomogeneous spaces*, Internat. Math. Res. Notices **15** (1997), 703–726.

[16] ———, *Accretive system Tb-theorems on nonhomogeneous spaces*, Duke Math. J. **113** (2002), no. 2, 259–312.

[17] ———, *The Tb-theorem on non-homogeneous spaces*, Acta Math. **190** (2003), no. 2, 151–239.

[18] ———, *Two weight estimate for the Hilbert transform and Corona decomposition for non-doubling measures* (2005), available at http://arxiv.org/abs/1003.1596.

[19] ———, *Two weight inequalities for individual Haar multipliers and other well localized operators*, Math. Res. Lett. **15** (2008), no. 3, 583–597.

[20] F. Nazarov and A. Volberg, *The Bellman function, the two-weight Hilbert transform, and embeddings of the model spaces K_θ*, J. Anal. Math. **87** (2002), 385–414. Dedicated to the memory of Thomas H. Wolff.

[21] Nikolai Nikolski and S. Treil, *Linear resolvent growth of rank one perturbation of a unitary operator does not imply its similarity to a normal operator*, J. Anal. Math. **87** (2002), 415–431. Dedicated to the memory of Thomas H. Wolff.

[22] Carlos Pérez, S. Treil, and A. Volberg, *On A_2 conjecture and Corona decomposition of weights* (2010), available at http://arxiv.org/abs/1006.2630.

[23] Stefanie Petermichl, *Dyadic shifts and a logarithmic estimate for Hankel operators with matrix symbol*, C. R. Acad. Sci. Paris Sér. I Math. **330** (2000), no. 6, 455–460 (English, with English and French summaries).MR1756958 (2000m:42016)

[24] Alexei Poltoratski and Donald Sarason, *Aleksandrov-Clark measures*, Recent advances in operator-related function theory, Contemp. Math., vol. 393, Amer. Math. Soc., Providence, RI, 2006, pp. 1–14.

[25] Maria Carmen Reguera and James Scurry, *On joint estimates for maximal functions and singular integrals in weighted spaces*, PAMS, to appear (2011), available at http://www.arxiv.org/abs/1109.2027.

[26] Donald Sarason, *Products of Toeplitz operators*, Linear and complex analysis. Problem book 3. Part I (V. P. Havin and N. K. Nikolski, eds.), Lecture Notes in Mathematics, vol. 1573, Springer-Verlag, Berlin, 1994, pp. 318-319.

[27] ———, *Exposed points in H^1. I*, The Gohberg anniversary collection, Vol. II (Calgary, AB, 1988), Oper. Theory Adv. Appl., vol. 41, Birkhäuser, Basel, 1989, pp. 485–496.

[28] Eric T. Sawyer, *A characterization of a two-weight norm inequality for maximal operators*, Studia Math. **75** (1982), no. 1, 1–11.

[29] ———, *A characterization of two weight norm inequalities for fractional and Poisson integrals*, Trans. Amer. Math. Soc. **308** (1988), no. 2, 533–545.

[30] E. Sawyer and R. L. Wheeden, *Weighted inequalities for fractional integrals on Euclidean and homogeneous spaces*, Amer. J. Math. **114** (1992), no. 4, 813–874.

[31] A. Volberg, *Calderón-Zygmund capacities and operators on nonhomogeneous spaces*, CBMS Regional Conference Series in Mathematics, vol. 100, Published for the Conference Board of the Mathematical Sciences, Washington, DC, 2003.

[32] Dechao Zheng, *The distribution function inequality and products of Toeplitz operators and Hankel operators*, J. Funct. Anal. **138** (1996), no. 2, 477–501.

School of Mathematics, Georgia Institute of Technology, Atlanta GA 30332, USA
E-mail address: lacey@math.gatech.edu